AFOSR 69-0082TR

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
UNITED STATES AIR FORCE
ARLINGTON, VA. 22209

WATERSHEDS AND INFORMATION FLOW*

ROWENA W. SWANSON

January 1969

*Prepared for presentation at the January 1969 meeting of the Boston Chapter of the Special Libraries Association entitled "Networks For Knowledge," and for publication in Special Libraries.

29

A: [illegible]

"Oh, no, not again," Paul moaned softly as the shiny metal filament behind the glass wall snapped.

"Tough luck," Jim commiserated beside him. "This means we've got budget for just one more try."

"Let's go to my office, get some coffee, and call Howie. It just could be that he'll be able to find something that can help us."

The two men left the testing lab. The bright morning sun and the warm browns and yellows of Paul's office soothed frayed nerves a little.

"Howie, Paul here. That last alloy we tried almost made it. I'm sure if we could get just a little more beryllium into it, it would far exceed what we need for the instrument. What are the chances you can locate some more information or someone who's had experience with this sort of thing?"

"I'll recheck the information files, Paul. Could be there's something in the pipeline. Also, I'll send out an S.O.S. to some of my colleagues in other lab info centers. We'll try to let you know by the afternoon."

Howard Blumberg shoved his chair to the typewriter console beside his desk. He was one of the lab's information troubleshooters. A metallurgist by training and experience, he had been attracted to the Bonhomme Institute because of its working environment and the caliber of its staff. He became impressed with a feature of its operations, a number of information services that provided useful backup to its engineers, researchers, and managers in the various specialty areas of its departments. He'd never thought of an information service as a vital part of a team before. He'd come to appreciate the information people at B.I., though. Sure, some of them were types he'd encountered before. Like Mrs. Winkle, trim but motherly, a stereotype of the librarian. And Claire and Louise, who cataloged and did some reference and charge-out work, though with the automatic system, people didn't have to handle a lot of mundane clerical chores any longer. Even before one got to Harry and Marge, though, one noticed a difference. Harry and Marge were information specialists. They read a lot, wrote abstracts that were circulated to people selectively according to their interests. But they did a lot more. They worked with several research groups, sat in on their seminars and briefings, and did a lot of leg work by telephone and typewriter console to locate information that could be useful to the research groups. Sometimes they'd also make surveys. The instituted was complimented more than once on these, because the surveys anticipated "hot" areas and saved people time, outside the Institute as well as in Institute labs, when they were ready to turn to these areas. Harry and Marge were a new breed, all right, but so were Mrs. Winkle and Claire and Louise. One got the feeling that they knew what you were talking about and why you needed what you asked for. They could even unscramble garbled recollections of books and papers and people, and it wa[illegible] a matter of pro[illegible] fessional pride to them to do all this Members of a team -- they sure were! Howard had gravitated to information work when the Institute received a contract for the quick study and development of an esoteric material that he had personally been keeping a file on. Amazing that this should have payed off on what he had often chuckled to himself and called a burrowing idiosyncrary.

Within half an hour, Howard knew that no useful new information had been received at the Institute's information centers that related to Paul's problem. He had also sent messages to two other centers, in California and Maryland, to recheck their files. A message went to a focal point in England, and one would be waiting for his friends in Japan when the time came round to opening that morning.

Howard stared out the window. "Dr. Lindstrom," he mused, thinking of the white-haired emeritus professor he'd met one evening during a "bull" session at the ASM meeting last year. Lindstrom had talked about several alloys he'd experimented with almost a lifetime ago in Europe. Hadn't he said he'd begun to look into these again here? A long-distance call to Texas.

"Paul, Howie here. A Dr. Lindstrom from Texas will be calling you. Sounds like he may have some information you can use."

"Thanks, buddy. The call's coming in right now on another line. Call you back later."

"Howie, Paul. That contact you made for me with Lindstrom might be it. He says he's gotten the properties you want, but only in the laboratory. We think we can put it on a production basis with a few processing changes. We're going to have a conference call this afternoon with Lindstrom. Some of our production men and lab men. Can you get in on it? There may be a few odds and ends we'll still need data for."

During the conference call, several processing bugs were discussed. Other groups at the Institute had already encountered several of them in connection with the development of other materials, as disclosed from a check of the information systems. The solution of one fabrication problem was noted in the reply received from England. The California center said it had had a request for similar information from a client several months ago. If the Institute got any information, would it send it on?

"Well, Lindy, that little trick of yours works. Guess this wraps it up." Weary, but beaming, the two men walked from the lab, leaving a few others from engineering and development to backslap each other and carry the work to its final stages of incorporation into a precision instrument.

[illegible] how thin [illegible] thread we have arrived at this point. I haven't told you yet why I was trying to develop this particular material. At a lecture many years ago, a surgeon described the properties of a material he wanted to have that I thought could be satisfied by this alloy. Do you suppose we could now talk to some people who might know whether this application is still needed? And could we call this the [illegible] alloy for the surgeon?"

Howie was called, and another search began, this time emphasizing surgical applications that once were no more than a vision in the mind of a man who thought to express it in a medical school lecture. How could he have known that a physicist attended that lecture that day who would find his way to the Bonhomme Institute in the United States and that this Institute would locate Dr. Falber at the Marburg Clinic, and, and, and....

Fact or fiction? The account above is fiction, but can one say it is not reflected by fact? There are few detailed records on how ideas are born, changed, embellished, disseminated, and on how substance is added to them until they become manifest in a theory or a product. One of the best records on the paths traversed by ideas from inception to fruition is the book written by John Jewkes and co-workers entitled <u>The Sources of Invention</u> (1). Jewkes realistically appraises the contexts in which industrial innovations have been achieved. The text also provides case histories for 51 major inventions. One cannot find, however, even from correspondence and histories, the full measure of happenings that either permitted an idea to be explored or to be abandoned (2).

The most fictitious aspect of the above account may be that pertaining to the information network and services structure that is described. Many information centers do exist in the United States and abroad (3). Information analysis and service operations, such as those attributed to the Bonhomme Institute, also exist. But coordination among these centers and services and an extensive program of interactive cooperation do not exist. Vocabularies differ. Internal methods of processing differ. Policies for external interaction differ. Existing facilities are plagued to various degrees by budget restrictions, staff limitations, acceptance by both management and clients, and uncertainty as

---

1. John Jewkes, David Sawers, and Richard Stillerman. <u>The Sources of Invention</u>. London: MacMillan & Co., Ltd., 1958.

2. See several accounts summarized in Rowena W. Swanson. <u>Information, An Exploitable Commodity</u>. Arlington, Va.: Air Force Office of Scientific Research, April 1968. AFOSR 68-0652; AD-677,197. Pages 4-12.

3. Descriptive information for a large number of information systems in the United States is contained in the publication <u>Nonconventional Scientific and Technical Information Systems in Current Use</u>. Washington, D.C.: National Science Foundation, Dec. 1966. NSF 66-24. Available from the U.S. Govt. Printing Office, $1.75.

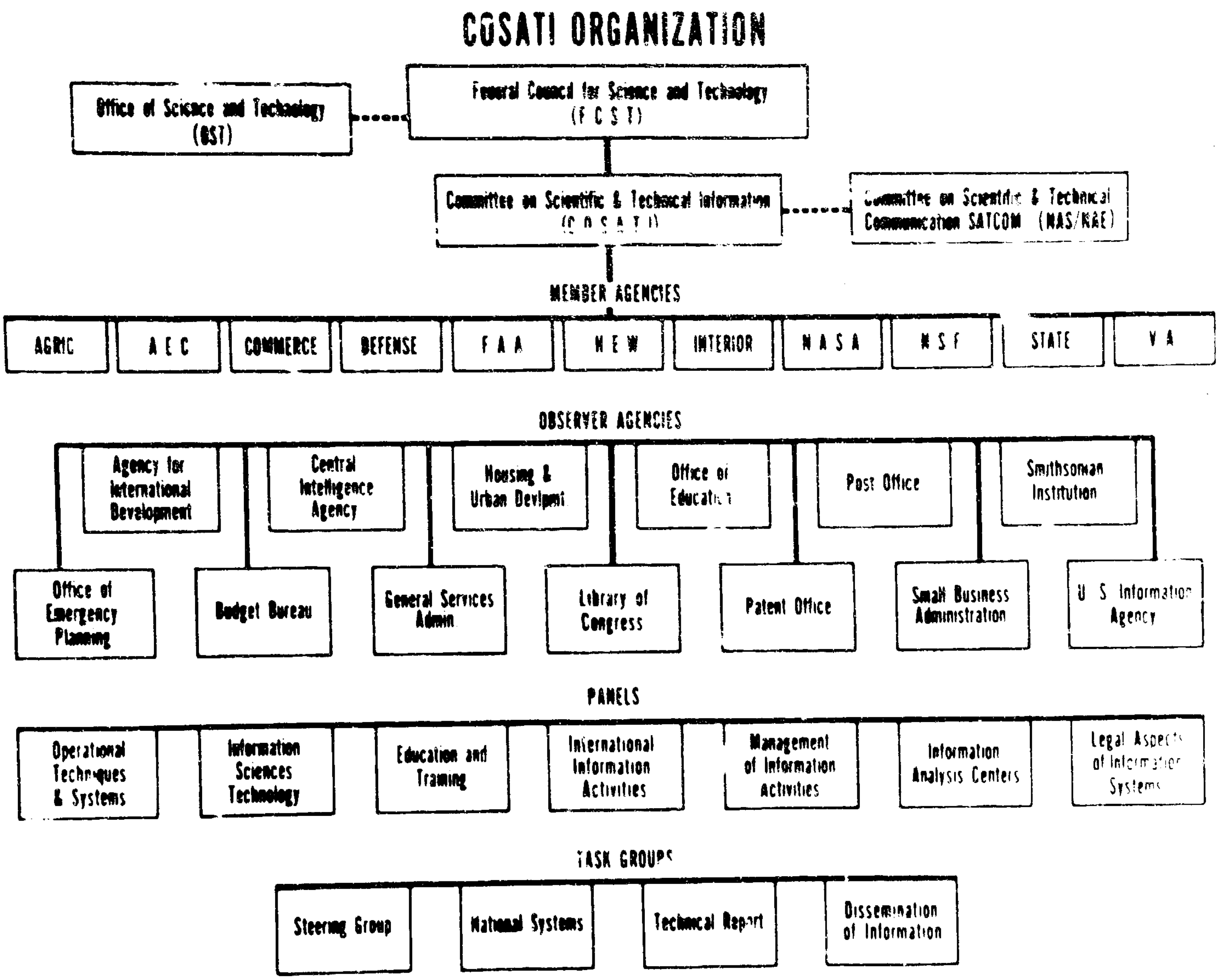

Figure 1. COSATI Organization Chart.

Reproduced in Reference 4, Vol. I, from Progress of the United States Government in Scientific and Technical Communication. Executive Office of the President, Federal Council for Science and Technology, 1966.

[illegible]

Watersheds for Information

Information is not like the weather. It is man-made and, increasingly, man has been doing more than merely talking about the problems it makes for him. Books and periodicals have been the usual vehicles for recording information. Libraries have been the usual means for storing these records according to some order so that they could be retrieved.

As long as the number of records on any given topic was small; as long as the growth rate was low, tractable, and relatively constant; as long as the number of people interested in the topic was small; as long as communication among them could be accomplished by correspondence and long time delays were tolerable, libraries were adequate to serve information needs. True, it would have been helpful to geneticists if Mendel's paper, buried in an obscure journal, could have been more widely publicized. Historians of science have traced similar instances. These are the isolated, dramatic examples of the inadequacies of libraries as storage and dissemination mechanisms. But are they so isolated? Might the growth of scientific progress, industrial development, and western civilization and expansion have been greater and its pattern even different with better methods of packaging and disseminating information?

The two World Wars of this century catapulted most of the major combatants into the present era of scientific exploration and technological development. Part of doing science has always been the publication of results. Though systems for storing and retrieving these results may not yet be generally accepted by the scientific community, or by anyone else, as part of doing science, the exponential growth of publication has forced the development of methods for storage and retrieval.

The largely unorganized, undirected proliferation of storage and retrieval activities during the past twenty years is, itself, a phenomenon of the present era. The United States has undertaken to apply some measure of scientific inquiry to the national-scale implications of this phenomenon through the Committee on Scientific and Technical Information (COSATI), established in 1962 by the Federal Council for Science and Technology. COSATI, in turn, formed a number of panels and task groups to study various information storage-and-retrieval problem areas (see Figure 1).

The COSATI Task Group on National System(s) for Scientific and Technical Information, chartered in 1964, has sponsored several studies of factors relevant to information system structures. One of these, conducted by Science Communication, Inc. (SCI), Washington, D. C. (now McLean, Va.), is entitled Study of Scientific and Technical Data Activities in the United States, issued in April 1968. In three volumes, this report both surveys a wide variety of scientific and technical data activities in industry, the professions, and government, and presents, with much background discussion, a plan for further study and implementation of national data system concepts (4).

The comprehensiveness of Volume II (bound in two reports) of the SCI study obviates the need to describe, here, various national and special information centers and systems, how they interact, and how they attempt to serve various communities of users (5). Rather, an attempt is made here to distill some of the essence of what the SCI investigators learned from the designers, operators, and users of the systems that bears on the questions: What do we do with what we've got? Where do we go from here?

At an over-all level, SCI has found that "existing organizational structures and human competencies are over-extended in terms of their abilities to effectively accommodate evolving data management and data handling needs. In addition, alleviation of this situation does not appear imminent, for only now are the individuals and organizations affected beginning to recognize the gravity of the situation. To date, attempts to improve data management and data handling operations have been rendered largely ineffectual due to crisis-type actions and piecemeal approaches to broad problems. Significant improvements cannot be expected until needs are better defined, organizational responsibilities are identified and accepted, and increased funds are made available to support the effort required to alleviate existing problems" (6).

Prevalent problem areas include those pertaining to personnel, finances, equipment, administration, input data, and coordination. The need for both systems- and subject-oriented personnel appears one of the most crucial. People are needed who combine talents of both types. Adequate training programs, both stop-gap and long range, have yet to be developed. High costs of equipment, computer services, and space for housing computer facilities are compounded by current budget squeezes. Opinions differ on

---

4. Study of Scientific and Technical Data Activities in the United States. Volume I - Plan For Study and Implementation of National Data System Concepts. Volume II, Parts A and B - Preliminary Census of Scientific and Technical Data Activities. Volume II, Part C - Preliminary Census of Scientific and Technical Data Activities. Final Report, ARPA Order No. 892, Contract F44620-67-C-0022, April 1968. AFOSR 68-1403, parts 1, 2, and 3; AD-670,606, AD-670,607, AD-670,608, respectively.

5. Many other papers and reports also contain descriptions of centers and systems. Some of these are cited in the "Selected Reading List" in Vol. I of the SCI report. References 2 and 3 above are also sources for descriptions. Systems being developed by libraries are summarized by the author in Move The Information ... A Kind of Missionary Spirit. Arlington, Va.: Air Force Office of Scientific Research, June 1967. AFOSR 67-1247; AD-657,794.

6. Reference 4, Vol. II, Parts A and B, p. 364.

future sources of funds, some suggesting the federal government, others state governments or service users. The unreliability of computer equipment for continuous on-line operations has necessitated backup systems and discouraged some experimentation. Few simple interpretive programs and languages exist that can be easily learned by users, thereby also slowing experimentation and direct interaction by non-programmer users and systems personnel. Few program and hardware configurations exist that enable a remote user, at his discretion, to work with files in a batch or computing mode or in conversational interaction. Many users need the capability of being able to work with a large variety of analog data and of being able to obtain transformations from analog to digital form. Many administrators who control budgets are not well informed on the value of systems and cost and personnel requirements. Many users are critical of the quality of the data in the systems. Lack of coordination of systems, sometimes even within a given institution, has produced duplication of effort, gaps in coverage, confusion to users, and skepticism that the morass can be resolved.

Approximately 100 issues were identified by SCI and are individually examined in the following five categories: (a) data management and handling system requirements, (b) data packaging, (c) data handling equipment, (d) personnel capabilities, and (e) institutional roles. For each grouping, panels of experts, totaling 300, provided their assessments. Discussions in the SCI study summarize these viewpoints. The comments in their entirety indicate, to SCI, "the complexity and limited level of knowledge which currently exists concerning requirements for data management on a national scale, and the data handling systems which would be responsive to these requirements" (7).

Though the import of the SCI study is that inadequate knowledge of user requirements and the reluctance of organizations and individuals to enter cooperative arrangements to alleviate common information problems are major deterrents, operational information systems and networks are practical possibilities. The study predicts: "Radical changes in data handling systems probably will not occur quickly simply because system designers do not know precisely the data service needs of scientists and technologists. In addition, even when service requirements are well defined, the system designer does not yet know how to effectively match data handling equipments and methods to data service requirements. Quite likely new data handling systems will be introduced in an evolutionary fashion with the computer first used as an aid for structuring, storing and formatting data for distribution in conventional forms. Knowledge gained from such experience will then be applied to implementation of more highly automated systems including query processing capabilities. Similarly computer techniques which are already widely used in design and other data manipulation operations at the work station of the scientist or technologist will continue to be refined and expanded in application. It, there-

7. Reference 4, Vol. I, p IV-1.

PHASE DESIGNATION	PHASE I	PHASE II	PHASE III	PHASE IV
PROGRAM ELEMENT	PLANNING		IMPLEMENTATION	
	NATIONAL DATA PROGRAM DEFINITION	NATIONAL DATA SYSTEM(s) CONCEPT DEFINITION	NATIONAL DATA SYSTEM DEVELOPMENT	NATIONAL DATA SYSTEM OPERATION
TIME SEQUENCE	1 YEAR	2 YEARS	3 YEARS	CONTINUOUS
RELATIVE LEVEL of EFFORT by FUNCTION				
CENTRALIZED PROGRAMMING*	1	>1	1	<1
PLANNING & COORDINATION	2	20	60	20
IMPLEMENTING & OPERATING	0	10	100	>500

*Arbitrarily designated as 1

Figure 2. Phases of SCI's Proposed National Data Program Plan Reference 4.

fore, does not appear unreasonable to anticipate a future merger of data handling systems to serve all of the scientist's or technologist's data handling needs, both archival and day-to-day manipulation. At least this possibility provides a future frame of reference which possibly can guide data handling system development efforts" (8).

The SCI study proposes a 4-phase plan for implementation over a six-year period (see Figure 2). The plan is consonant with COSATI's recognition that there be no disruption of existing information operations and that account be taken of the widely differing capabilities of existing systems and realities of funding, long-established practices, rapid changes in information technology, and differing needs of the various segments of the user communities. Moreover, the Government can only encourage, not direct, private activities, to join in a national undertaking. The plan is called a "preliminary blueprint" for establishing a National Scientific and Technical Data Program.

Under the plan, SCI envisions a one-year Phase I for organization that would include the formation of a Scientific Data Program Office in the National Science Foundation, a Technical Data Program Office in the Department of Commerce, a Data Systems Technical Information Center to support the program, and a National Advisory Council for Scientific and Technical Data whose panels would provide input for the program plan. A two-year Phase II is specified for critically reviewing the requirements for data management and data handling systems at local and national levels, and for conducting development and prototype tests to determine the adequacy of equipment and methods for meeting management and handling requirements. A three-year Phase III contemplates the development and testing of methods and facilities for serving the data handling needs of specific communities. It anticipates that several systems will be under development concurrently and will differ substantially as to structure and function, some being centralized and others decentralized dependent on the needs of the community to be served. Standardization efforts are expected during this period so that the various systems can be compatible. Though this development phase is expected to be evolutionary for specific scientific and technical communities, increasing integration within federal agencies is foreseen for general-purpose data collection. The operational Phase IV should increasingly involve non-government organizations that are expected to voluntarily participate in the program.

Whether or not SCI's plan is adopted in whole or in part, the factual portion of its survey should be highly beneficial to all individuals and groups concerned with any aspect of information system design, operation, or use. This author concurs with SCI's conclusions, considers them factual, and believes that SCI's recommendations are realistic. (See the Appendix to this paper which lists the major conclusions and recommendations contained in the SCI report.)

---

8. Reference 4, Vol. I, p. IV-4.

Information has been called a national and international resource by a U.S. Senate Committee (9) and others. It is a commodity that is expensive to private and public groups to produce. It has become an expensive commodity to package for use. It has many different forms and is needed in many different ways by many different people, and now, also, by many different machine-based systems. Sound methods of handling information for the various people and purposes it ought to serve are urgent and critical for the effective functioning of military, scientific, industrial, and social organizations. The all-pervasive nature of the commodity and the need and the long-term deficit spending required for integrated program development make the federal government the focal point for the national effort and impose pressure on the government to establish and put a broad-scope plan into operation. It seems inevitable that some version of the SCI projections will begin to be implemented during the next few years. Ultimate returns on [illegible] will depend [illegible] investments of people, time, and money, and understanding cooperation among all the participants, throughout the program development.

Information System Networks

Fortunately, existing organizations have accumulated much valuable experience that should be useful to national-scale program planners. One of the most productive of these, [illegible] the Office of Standard Reference Data [illegible] are, therefore, reviewed here for their [illegible] (10).

A June 1963 [illegible] and Technology, Executive Office of the President, [illegible] Bureau of Standards (NBS) "to pro[illegible] coordination and direction [illegible] quality standards, and also to [illegible] of the system at the level of the [illegible] Standard Reference [illegible] standards. The activities that [illegible] the National Standard Reference [illegible]

[illegible] the operation of [illegible] information [illegible]

---

9. [illegible] Report of the Committee on [illegible] United States Senate made by its Subcommittee [illegible] Organizations, U.S. Senate, 89th Cong., [illegible] 1965, Senate Rept. No. 369.

10. This review is based on the following report: Edward L. Brady, ed. *Status Report, National Standard Reference Data System, April 1968.* Washington, D.C.: National Bureau of Standards, June 1968. NBS Technical Note [illegible] 70 p.

In many instances, the OSRD has monitoring responsibility but not direct operating responsibility for data centers within the NSRDS. Many data compilation projects were started prior to 1963 by federal agencies and private groups to meet various mission objectives. They continue to receive funds and direction from these sponsors. Other projects are wholly sponsored by the OSRD, the work being performed in university or industry laboratories. Still other projects are joint ventures with NBS experimental divisions which conduct their own research and data compilation.

Coordination is accomplished through a variety of conferences and by correspondence and publication exchanges. Intermittently, OSRD convenes meetings of data project operators. Panels of specialists called by its Advisory Committee consult on data problems. OSRD maintains contact with such international groups as the Committee on Data for Science and Technology (CODATA) of the International Council of Scientific Unions [illegible] with comparable coordination organizations in other countries. It publishes a monthly newsletter, NSRDS News, that is both a section in the NBS Technical News Bulletin and a separate bulletin mailed to 3200 recipients. It disseminates information on reports and compilations produced by member centers and projects. In cooperation with the journal, Industrial Research, it conducted a national survey on data needs to inform itself and as a service to member projects. It has helped establish operational criteria for projects with respect to comprehensiveness of literature reviews, critical evaluation procedures, and continuity requirements. With the aid of its Advisory Committee, it has clarified a basic question concerning the terms "standard reference data" and "reference data."

An in-house research group of three professionals is currently working on guidelines for the preparation of general-purpose computer programs for data storage and retrieval, for file manipulation, and for [illegible] preparation, editing, and printing. The group is developing [illegible] for remote access to a computer-based file that may be attractive to university and commercial centers. The plan is to organize subject matter for applicability to chemistry and physics instruction. [illegible] program, called EDPAC, was written for text preparation [illegible] in ASA Standard Fortran IV, making it operable on a variety of machines. [illegible] in the EDPAC series have been in use at NBS [illegible] programs that convert existing data tapes for automatic typesetting have produced several NSRDS publications. Current programs [illegible] operations, but have demonstrated suitability of the approach to data tables of complex structure. The group is gaining useful experience on hardware for data processing applications and on needed hardware configurations. It can, thereby, assist member projects with mechanization problems.

Figure 3 depicts the interrelationships among one community of special information centers in the NSRDS, namely those that handle nuclear data. Dashed lines in the figure indicate advisory roles or cooperative

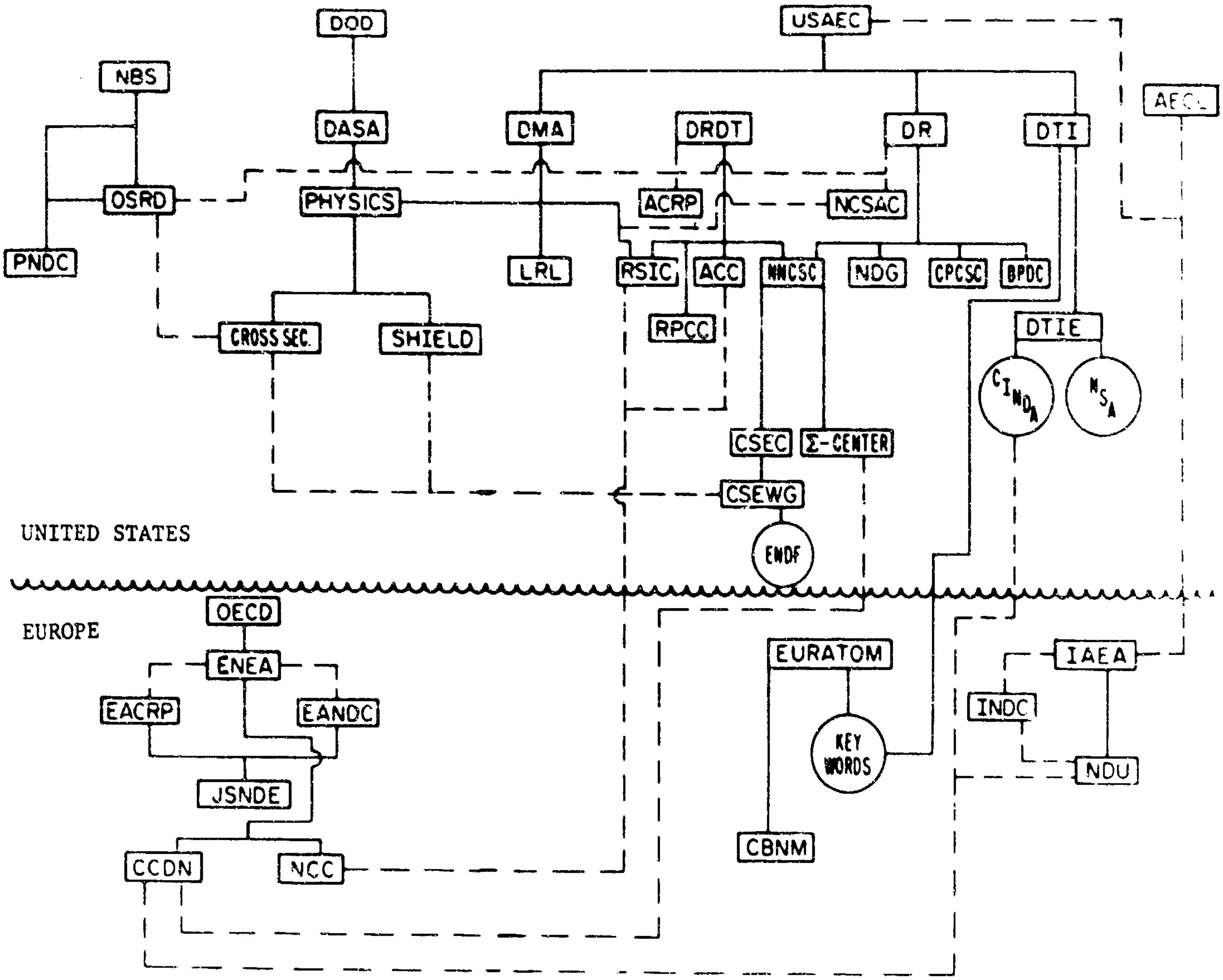

Figure 3. Nuclear Data Compilation and Evaluation Systems Community
Reference 10.

activities. Noteworthy is the degree of international collaboration through both formal agreements and informal arrangements. The Ispra Code Center (NCC), the Argonne Code Center (ACC), and the Radiation Shielding Information Center (RSIC), for example, have been exchanging computer programs for several years. CINDA, an annotated bibliography of neutron cross-section information, is published by the Oak Ridge Division of Technical Information Extension (DTIE) from input provided by, among others, the Neutron Data Compilation Center (NDCC) in France, the Nuclear Data Unit of the International Atomic Energy Agency (IAEA) in Vienna, and the U.S.S.R. Nuclear Data Information Center. A high degree of information exchange exists in this community among data generators and users and information system personnel. Though the performance of this community is, no doubt, based on its being one of the oldest (25 years), best integrated, best sponsored, and technologically most advanced networks, its existence reinforces projections for the successful realization of other complex networks.

Information Flow

Too frequently overlooked in discussions of information systems is the most important element of all, the man in the system. Even in systems where computers "talk" to computers, a man at some point triggered the "conversation" for an ultimately man-desired purpose. The man in the system is more than a statistic, either as a member of the system staff or as a producer or user of information. Awareness of the need to understand how man cognitively processes information as well as his characteristics as a communicator and his motivations as a producer and generator is only gradually entering the information systems context, though these have long been subjects of investigation of behavioral and social scientists.

A beginning on studies of the man in the system is evident in literature on information flow and transfer, user needs, and evaluations of systems. The author has reviewed some of this work elsewhere (11). Notable among recent investigations is the comprehensive Evaluation of the Medlars Demand Search Service conducted by F. W. Lancaster (12). Findings can be related to the performance of a large variety of types of people who directly or indirectly influence an information system. For example, lack of consistency and the omission of terms by human indexers suggest to Lancaster tools to assist them and possible modifications in methods of

11. Information flow studies are considered in Reference 2, user needs in Reference 5.

12. F. W. Lancaster. Evaluation of the Medlars Demand Search Service. Bethesda, Md.: National Library of Medicine, Jan. 1968. See also: Madeline M. Henderson. Evaluation of Information Systems: A Selected Bibliography with Informative Abstracts. Washington, D.C.: National Bureau of Standards, Dec. 1967. NBS Technical Note 297. Available from U.S. Govt. Printing Office, $1.00.

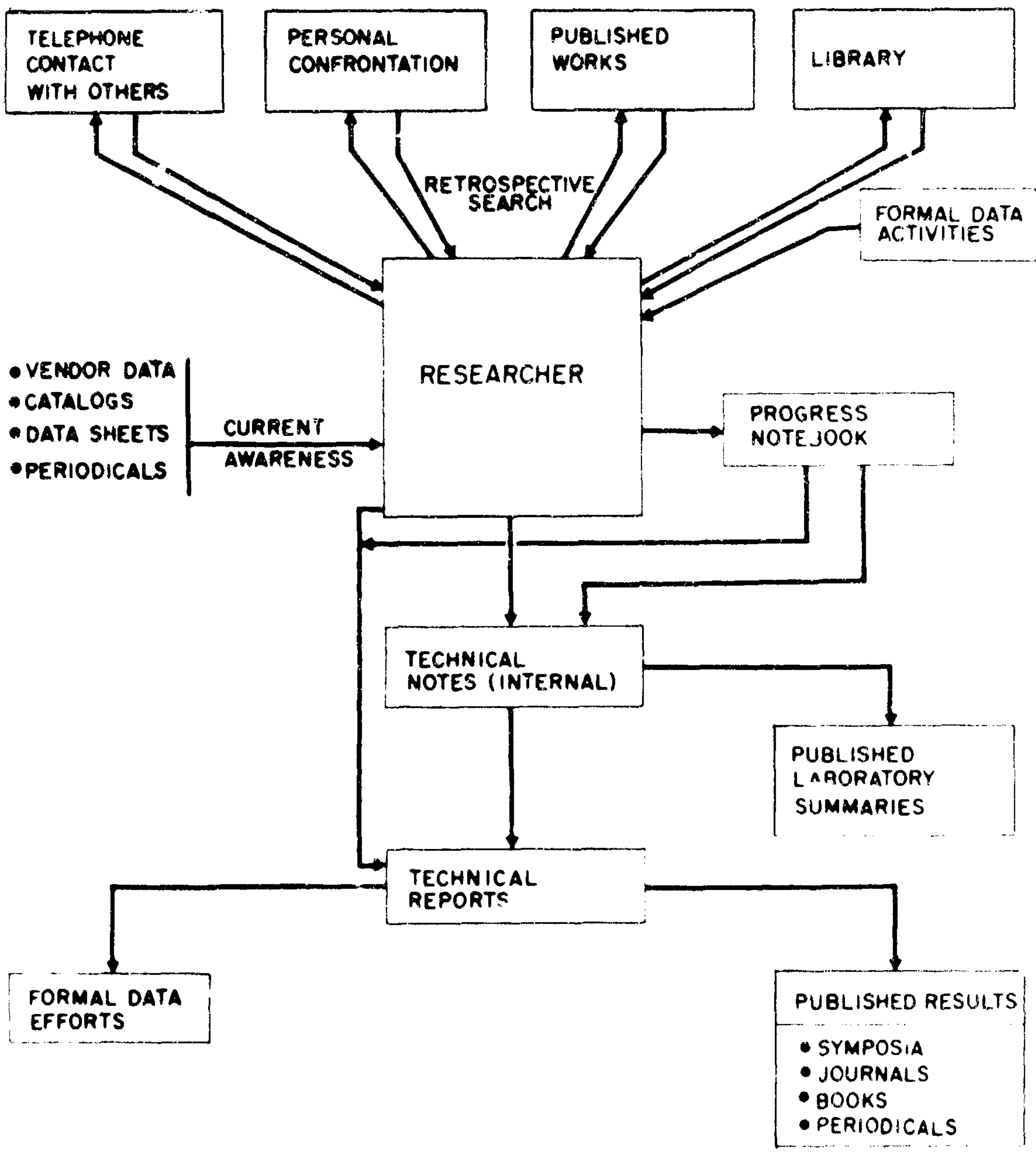

Figure 4. Information Flow Environment in Research Communities Reference 4.

vocabulary development. Comprehension failures were observed at the user-information service personnel interface. The best request statements that clearly reflected user needs were those written by the user in his own, natural language, narrative terms, not those that were paraphrased and recorded by system personnel or translated by them or by the users into their concepts of the system vocabulary. On the other hand, education of the user seems necessary to condition him to be precise in stating his need with, preferably, his tolerances and some background on what his need is for. A summary, however, cannot do justice to Lancaster's detailed account and analysis.

The literature on people as information transfer agents or "brokers" mediating between information systems and users will probably be enriched soon by publications sponsored by the Office of State Technical Services (OSTS), U.S. Department of Commerce. The OSTS is charged, under the State Technical Services Act of 1965, with the administration of a national State and interstate technical services program that promotes the application of research results in business, commerce, and industry. OSTS projects are coping with problems people encounter in trying to discover user needs and match available knowledge with individuals who need it. One hopes that some OSTS contractors are considering the behavioral aspects of the information flow and transfer process.

Several charts in the Science Communication survey provide self-explanatory illustrations of the information-rich environments people work in (see Figures 4, 5, 6, and 7). These bear evidence that "keeping up with the literature" is an impossible euphemism. Experiments involving eye and ear inputs and vocal or pointing outputs give values for human information processing rates of 6 to 45 bits/sec., even though the ear and the eye, respectively, have channel capacities of 8,000-10,000 bits/sec. and 5,000,000 bits/sec. (13). It has been suggested that "a theory of memory that is based on mechanisms that store events is not only uneconomical bordering on the impossible, but also is incapable of explaining the most primitive types of behavior in living organisms that show one or another form of retention" (14). A reasonable postulate is that humans accomplish information processing through powerful methods of classification and abstraction. An upper bound of 10^{47} bits/g/sec. has been computed for the transmission capability of matter. Because of this limit, Ross Ashby negates hypotheses that man's brain abstracts through sequential operations that use primary sense data as the initial input (15). He also cautions that this limit imposes restrictions on realizable net-

---

13. James L. Flanagan. Speech Analysis, Synthesis and Perception. Berlin: Springer-Verlag, 1965.

14. Heinz von Foerster. "Computation in Neural Nets." Currents in Modern Biology, vol. 1 (1967) 47-93, at 91-92.

15. W. Ross Ashby. "Mathematical Models and Computer Analysis of the Function of the Central Nervous System." Annual Review of Physiology, vol. 28 (1966) 89-106.

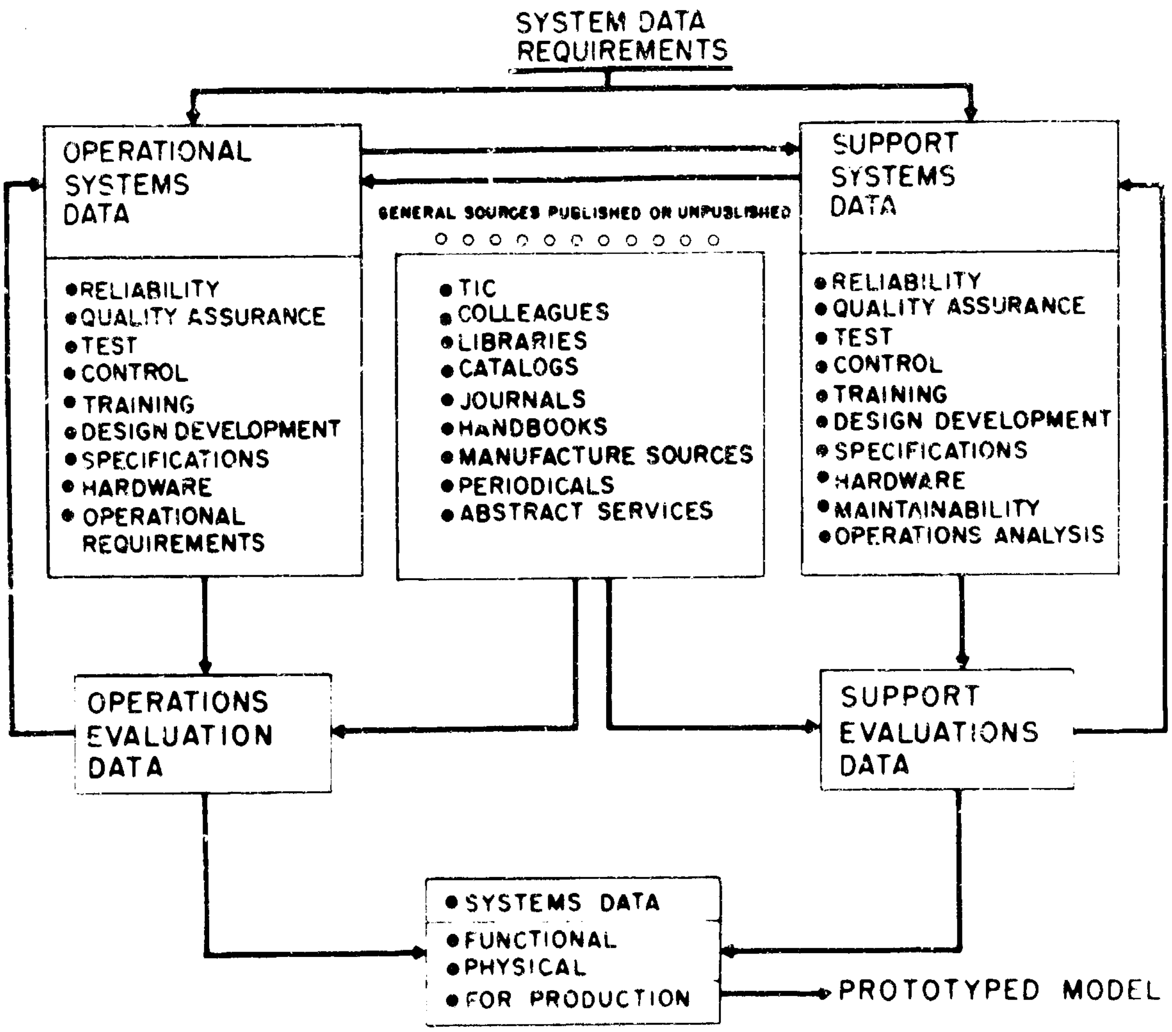

Figure 5. Information Flow Environment in Technology and Development Communities

Reference 4.

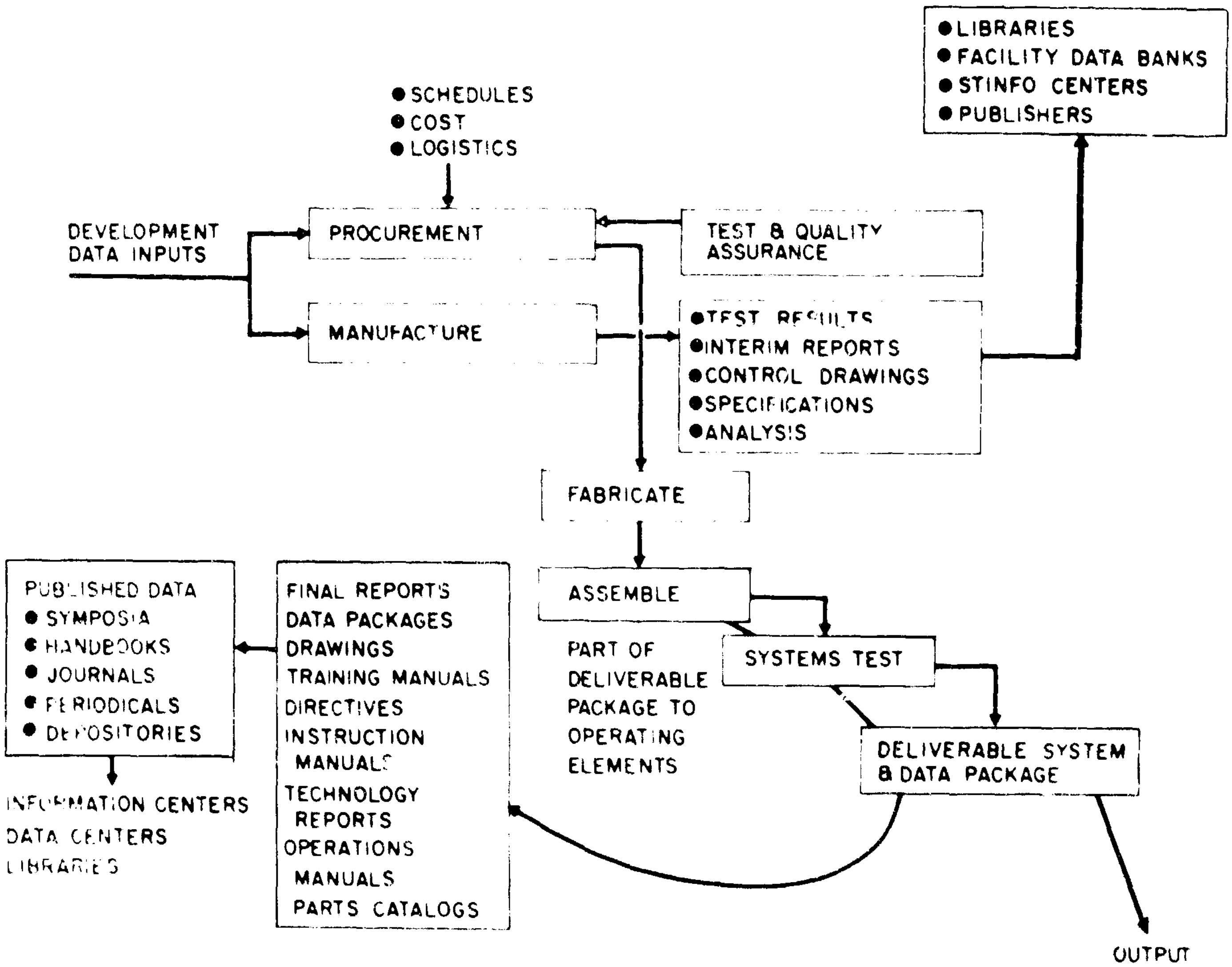

Figure 6. Information Flow Environment in Production Communities Reference 4.

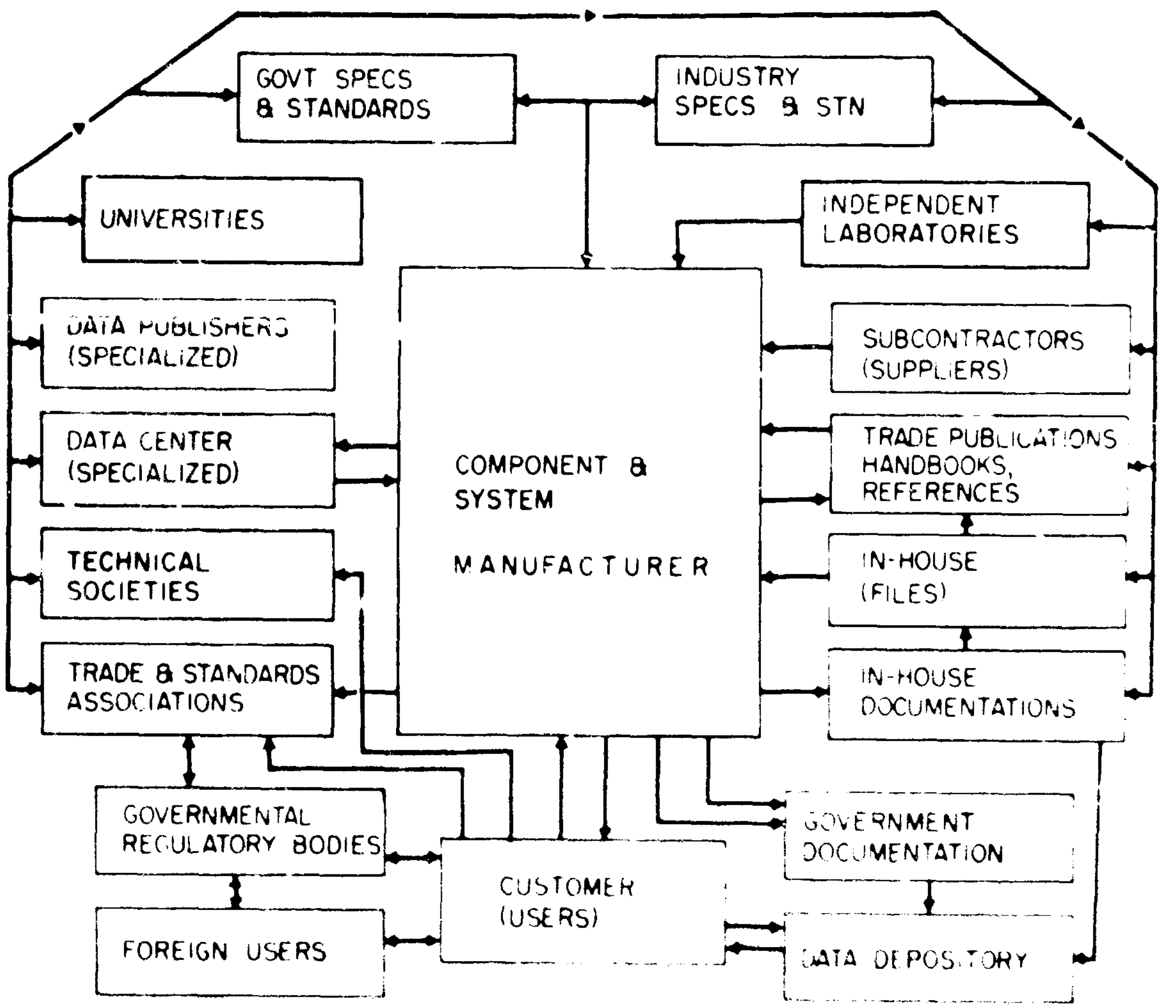

Figure 7. Information Flow Environment in R & D Communities (Leading to Production) Reference 4.

works of mechanical 'brains.' He exemplifies the seriousness of exponential growth with an illustration in which a difference of 10 from 10^{10} to 10^{20} search operations is the difference between a computation requiring 3 million years and one performable in a morning.

These inroads toward an understanding of physical, physiological, and behavioral potentials and limitations are little more than surface scratches. More studies are needed that are both systematic and directly related to the goals of coping with man's information environments and of adapting systems to men and men to systems. Pencil-and-paper, "demonstration," and "feasibility" studies should no longer be ends in themselves, but steps to applications for operating systems. It is not sufficient that investigations be made solely of existing conditions, as many of the information flow studies have done. These should be recognized by investigators and sponsors alike as only providing base lines against which tests of realistic duration and scope must be made of the effects of modifications and changes. The SCI survey, for example, discloses that volume needs vary with categories of data and with types of data processing (see Figures 8 and 9). How might quantification of these loadings simplify the data processing requirements of some systems? Might this stratification be useful in designing efficient data compression schemes? What flexibilities can be built into operating systems like Medlars that could permit the study of small or appreciable changes in vocabulary structure, computer programs, and duties of systems personnel? What mathematical modeling is being done to avoid large-number pitfalls in complex information processing?

A major distinction between information systems and conventional libraries is, or should be, the dynamic nature both of its services and its structure. Whereas libraries are passive repositories, information systems are supposed to be active interfaces between producers and consumers (producers, of course, also being consumers, and vice versa). Flow and evaluation studies have shown that systems have not been accorded the user acceptance they expected for a variety of reasons, many of which can be traced to characteristics of human behavior. Whereas library operations can remain largely the exclusive province of librarians, no such uniqueness adheres to information system operations. The entity "information system" and, even more so, the entity "information systems network," require that these be common meeting grounds for the many different types of people who are integral parts of the producer--system (network)--user loop. Only through interaction can these be information system operations. Time has been needed to make interdependencies obvious. Now is the time to share knowledge, pool talents, and engage in innovative thinking if information is to be what it is intended to be -- a tool in the service of mankind.

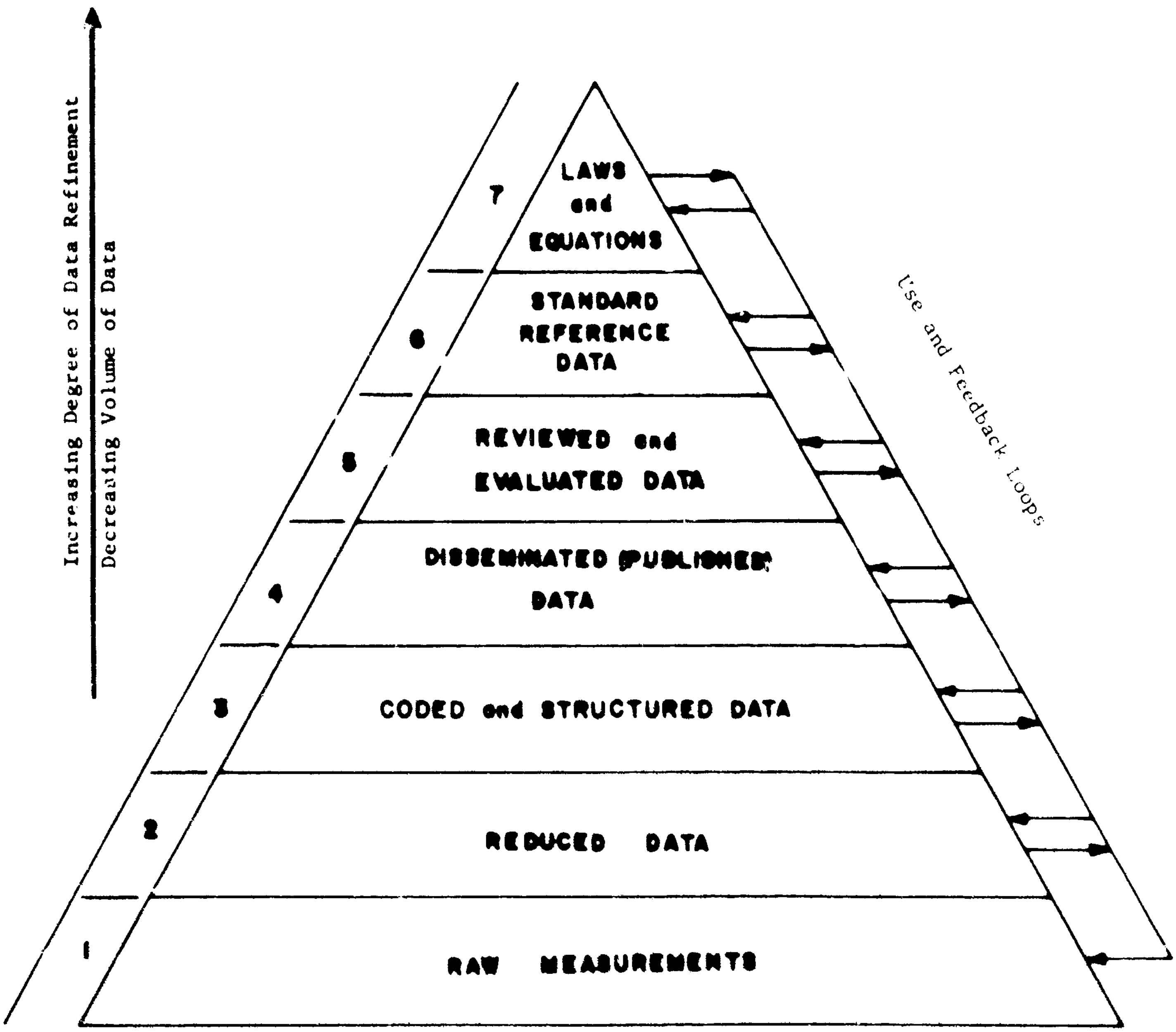

Figure 8. Patterns of Use of Categories of Data
Reference 4.

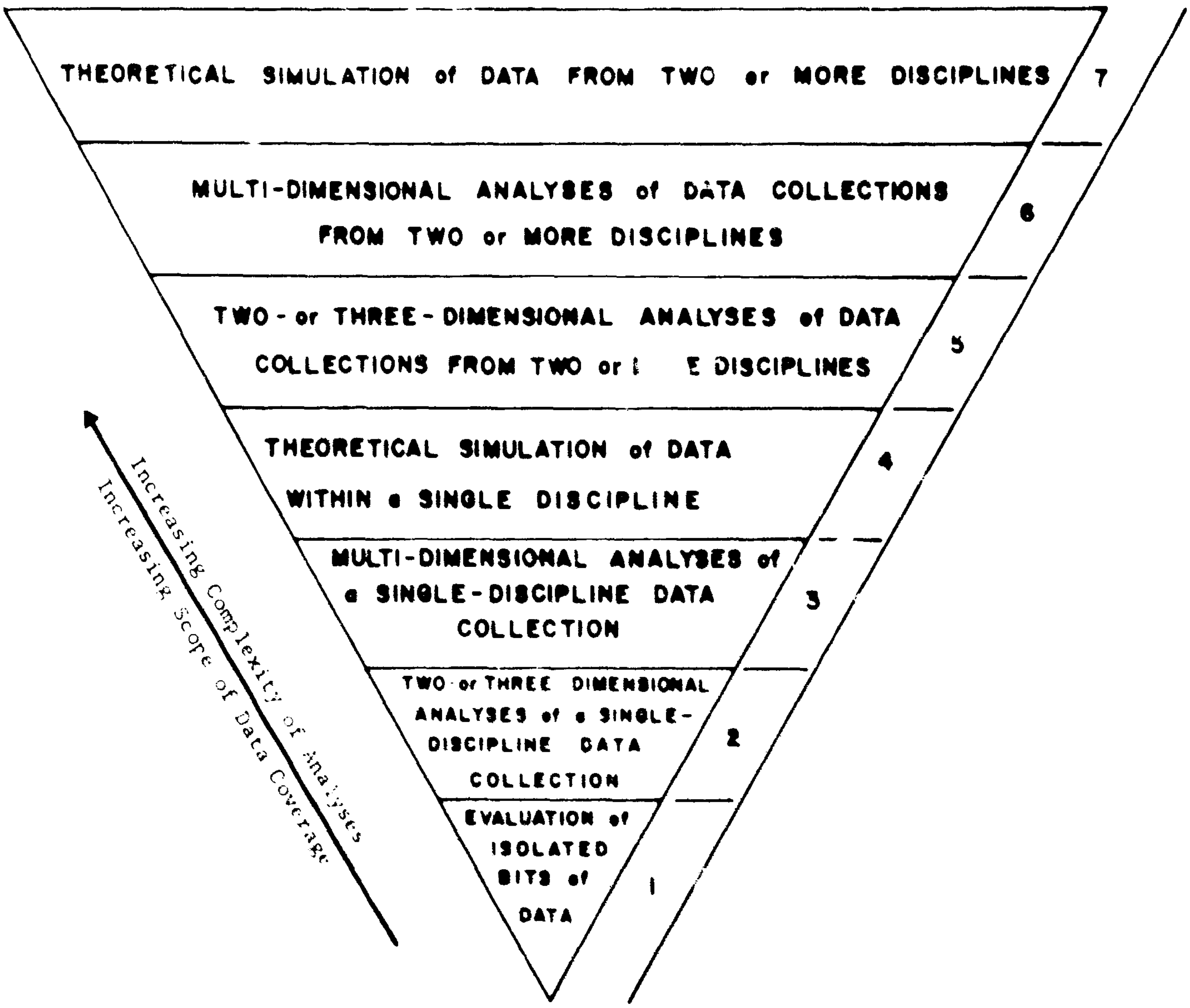

Figure 9. Levels of Interaction with Data (Types of Data Processing) Reference 4.

APPENDIX

MAJOR CONCLUSIONS AND RECOMMENDATIONS

Contained In

STUDY OF SCIENTIFIC AND TECHNICAL DATA ACTIVITIES IN THE UNITED STATES
Science Communication, Inc., Washington, D. C. (now, McLean, Virginia)
Final Report, Contract F44620-67-C-0022; 30 April 1968; AFOSR 68-1403, pt. 1; AD 670,606.

A. SCIENTIFIC AND TECHNICAL DATA - PERSPECTIVES AND POLICY IMPLICATIONS

1. *CONCLUSION:* The utility of our national scientific and technical data resource can be substantially increased by improved management.

RECOMMENDATION: The Executive Office of the President should issue a policy statement establishing the objectives of a national program to improve the management of scientific and technical data activities within government, the professions, and industry.

2. *CONCLUSION:* No effective means currently exists for coordinating and integrating the data management and data handling activities of the governmental, professional, and industrial sectors of science and technology.

RECOMMENDATION: A National Advisory Council for Scientific and Technical Data should be established.

3. *CONCLUSION:* Scientific and technical data and data activities are exceedingly complex: national data programs and system development efforts must be capable of effectively recognizing and accommodating this complexity.

RECOMMENDATION: National data programs and related policies must be implemented with due consideration of the diverse types of data activities which are conducted as an integral and vital part of science and technology.

4. *CONCLUSION:* The full utility of scientific and technical data is not currently realized under existing data management and data handling policies.

RECOMMENDATION: Each scientific or technical community, including mission-oriented agencies, should reappraise its current procedures for managing and handling scientific and technical data, especially in regard to their adequacy for conservation of the data as a costly and potentially reusable resource.

5. *CONCLUSION:* There is inadequate knowledge concerning the nature (quantity, quality, location, ownership, usefulness, etc.) of existing scientific and technical data to permit optimum design of national data management programs or data handling systems.

RECOMMENDATION: A National Index of Scientific and Technical Data should be developed. Such an index is essential if data management is to be planned on a systematic basis. Also, such an index would be immediately useful to scientists and technologists who currently expend as much as 30% of their working hours searching for data required to perform their job.

6. CONCLUSION: Federal policy relative to scientific and technical data management must recognize and facilitate maximum use of the existing scientific and technical data resource.

RECOMMENDATION: The Federal Government should establish a policy to encourage the accessibility of scientific and technical data to as many potential users as possible.

7. CONCLUSION: As data handling becomes increasingly automated, the need for standardization of data handling methods will become increasingly important to the National Scientific and Technical Data Program.

RECOMMENDATION: The Federal Government should take action to assure development and application of standardized methods of handling basic scientific data, especially those automated methods broadly applicable to data systems in more than one field of research.

8. CONCLUSION: The diverse connotations assigned by different communities, organizations, and individuals to scientific and technical data, data artifacts, and data management and handling efforts constitute severe barriers to systematic planning and evaluation.

RECOMMENDATION: The Committee on Scientific and Technical Information (COSATI) should promulgate a set of definitions which delineate an internally consistent set of terms covering scientific and technical data activities.

9. CONCLUSION: Just as science is international, scientific and technical data activity is often international in scope.

RECOMMENDATION: Offices in the Federal Government designated as responsible for representing U.S. interests in the area of international data activities should be strengthened not only to permit them to better represent U.S. interests, but also to enable them to establish better communications and working relationships with on-going activities in the U.S.

B. CURRENT ISSUES AND PROBLEMS - NATURE AND POSSIBLE RESOLUTIONS

1. CONCLUSION: The inadequacy of classifical methods for structuring and communicating scientific and technical data in current working contexts has created unnecessary apprehension.

RECOMMENDATION: The currently evolving expressions of need for large-scale scientific and technical data handling systems should be viewed as a response to opportunity, not an act of desperation to avoid inundation by the flood of data.

2. CONCLUSION: Current research and development administration, especially within Federally sponsored programs, frequently gives preferential consideration to research and development to generate scientific and technical data over activities directed to assembly, evaluation, and application of existing data.

RECOMMENDATION: Each Federal research and development program should be required to allocate a minimum percentage of its budget to husbandry and conservation of the scientific or technical data generated by the program.

3. *CONCLUSION:* Although essentially the same problems are observed in data handling activities in the different fields of science and technology, no mechanism now exists for the coordination of efforts toward solution of these problems.

RECOMMENDATION: The Federal Government should establish an information center to serve as a depository and dissemination agency for information dealing with design, development, operation and management of scientific and technical data systems.

4. *CONCLUSION:* Current data service requirements are largely undefined.

RECOMMENDATION: Existing data service centers such as the National Oceanographic Data Center and National Space Sciences Data Center, and new prototype data resource centers, should be used as test beds to study data service needs.

5. *CONCLUSION:* Current scientific and technical handling practices do not fully employ available technologies.

RECOMMENDATION: The Federal Government, professional societies, trade associations, commercial publishers, and other collectors and disseminators of scientific and technical data should explore means of applying modern technology for more effective assembly and dissemination of scientific and technical data.

6. *CONCLUSION:* The lag time between data generation and dissemination using traditional publications is frequently from two to five years.

RECOMMENDATION: Programs should be developed to more directly couple experimentation, tests, etc., with data systems.

7. *CONCLUSION:* Although the total investment applied to generation of data concerning products and processes far exceeds that applied to generation of basic scientific data, inadequate effort is expended by the Federal Government to organize this data for secondary uses.

RECOMMENDATION: Current efforts, such as the Department of Defense Engineering File, should be substantially accelerated, and other equipment development agencies without such systems should initiate study of their feasibility.

C. SYSTEMS DEVELOPMENT - REQUIREMENTS AND IMPLEMENTATION CONCEPTS

1. *CONCLUSION:* It cannot be expected that existing groups will cooperate in efforts to develop national systems where the purpose is intangible or Federal domination might restrict the legitimate freedoms of scientific groups, commercial firms, etc.

RECOMMENDATION: A National Scientific and Technical Data Program must be planned and administered in a manner which accommodates the interests and capabilities of diverse groups and organizations.

2. *CONCLUSION:* The inability to define a single system structure responsive to all data management and data handling requirements does not constitute a valid justification for delaying consideration of new or improved data systems.

RECOMMENDATION: The present should be recognized as a timely point for initiation of national systems planning and development efforts.

3. *CONCLUSION:* Data management is in a state of transition which is being driven largely by the introduction of computer and other improved data handling methods.

RECOMMENDATION: In the near future, efforts at the national level should be directed toward the development and test of systems or tools to facilitate better data management.

4. *CONCLUSION:* The most valid requirements for development of national scale data handling systems exist for systems operating within scientific and technical communities rather than between communities.

RECOMMENDATION: National data system development efforts should be focused on individual communities.

5. *CONCLUSION:* Data handling systems are tools to facilitate data management.

RECOMMENDATION: The total data management requirements of a community to be served should be examined prior to implementation of data handling systems as part of the National Scientific and Technical Data Program.

6. *CONCLUSION:* National scientific and technical data system development efforts must consider not only the scientific or technical field to be served by the system, but also the specific type or phase of activity to be served.

RECOMMENDATION: Federal support of national scientific and technical data programs and systems should be pro-rated according to the type of data activity served and the stage of the data program or system.

7. *CONCLUSION:* To be effective, data service operations must be complementary to the normal work routines of the scientist or technologist.

RECOMMENDATION: A part of the National Scientific and Technical Data Program should be the development of integrate [illegible] resource and service centers.

8. *CONCLUSION:* Data systems complement other information systems; however, it is short-sighted to view data systems as simple extensions of document handling systems.

RECOMMENDATION: Data management and handling systems in their ultimate form should be viewed as providing a capability for a totally new level of interaction between the scientist or technologist and the accumulated data resource.

D. SYSTEMS CAPABILITIES - ASSESSMENTS AND REMEDIAL ACTIONS

1. CONCLUSION: Current research and development directed specifically to study of critical factors important to develop large-scale scientific and technical data handling systems is totally inadequate.

RECOMMENDATION: The Federal Government should budget at least one-tenth of one percent of its total annual expenditure on research and development for research on techniques and procedures for managing and handling scientific and technical data.

2. CONCLUSION: Current personnel and institutional capabilities are not adequate to support a crash program to develop a national scope scientific and technical data handling system.

RECOMMENDATION: Information and data managers should be developed from two sources - one is the current population of working scientists, engineers, data processing specialists, etc.; the second is the current and future population of students in colleges and universities.

3. CONCLUSION: Although data switching networks and computers are frequently mentioned in juxtaposition to one another, automated data service networks, for all practical purposes, do not currently exist within science and technology.

RECOMMENDATION: Appropriate organizations should test the effectiveness of centrally supported, decentralized data resource centers as an alternative systems concept to data switching networks.

4. CONCLUSION: There is an almost complete absence of criteria for the evaluation of the economic performance of current data handling efforts.

RECOMMENDATION: Cost-effectiveness should not be the principal criterion to determine whether or not efforts should be initiated to explore the feasibility of improved data handling systems.

Unclassified

Security Classification

DOCUMENT ...OL DATA - R & D

Security classification of title, body of abstract ...indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Air Force Office of Scientific Research Directorate of Information Sciences Arlington, Virginia 22209	2a. REPORT SECURITY CLASSIFICATION unclassified 2b. GROUP
3. REPORT TITLE **WATERSHEDS AND INFORMATION FLOW**	
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) scientific; ; interim	
5. AUTHOR(S) (First name, middle initial, last name) Rowena W. Swanson	

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
Jan. 1969	26	12

8a. CONTRACT OR GRANT NO. n/a b. PROJECT NO. 9769-01 c. 61445O1F d. 681304	9a. ORIGINATOR'S REPORT NUMBER(S) AFOSR 69- 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFOSR 69-0082TR
10. DISTRIBUTION STATEMENT Distribution of this document is unlimited.	
11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY See 1 above.

13. ABSTRACT

This report discusses a plan for a national information system. It reviews the findings of a survey of scientific and technical data activities in the United States. Comments are included on prevalent problem areas involved in system and network construction that concern personnel, finances, equipment, administration, input data, and coordination. Organizational structures and human competencies appear over-extended in terms of their abilities to effectively accommodate evolving data management and data handling needs. Alleviation of the situation does not appear imminent, since the individuals and organizations affected are only beginning to recognize its gravity. Characteristics of the NBS National Standard Reference Data System are described in exemplification of an information system network format. Information flow is considered with respect to physical, physiological, and behavioral potentials and limitations. The successful realization of complex information networks is considered possible.

DD FORM 1473

Unclassified

Security Classification

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Information systems						
Information system networks						
Information flow						
Evaluation						
Human information processing						
National Standard Reference Data System						
Data processing						
Information transfer						